Agreement between routine and research measurement of infant height and weight

M Bryant,1,2 G Santorelli,2 L Fairley,2 E S Petherick,2 R Bhopal,3 D A Lawlor,4,5 K Tilling,4,5 L D Howe,4,5 D Farrar,2 N Cameron,5 M Mohammed,7 J Wright,2 the Born in Bradford Childhood Obesity Scientific Group

ABSTRACT
In many countries, routine data relating to growth of infants are collected as a means of tracking health and illness up to school age. These have potential to be used in research. For health monitoring and research, data should be accurate and reliable. This study aimed to determine the agreement between length/height and weight measurements from routine infant records and researcher-collected data.

Methods Height/length and weight at ages 6, 12, and 24 months from the longitudinal UK birth cohort (born in Bradford; n=836–1280) were compared with routine data collected by health visitors within 2 months of the research data (n=104–573 for different comparisons). Data were age adjusted and compared using Bland Altman plots.

Results There was agreement between data sources, albeit weaker for height than for weight. Routine data tended to underestimate length/height at 6 months (0.5 cm (95% CI −4.0 to 4.9)) and overestimate it at 12 (−0.3 cm (95% CI −0.5 to 4.0)) and 24 months (0.3 cm (95% CI −4.0 to 3.4)). Routine data slightly overestimated weight at all three ages (range −0.04 to 0.04 kg (95% CI −0.12 to 0.9) to −0.04 to 0.04 (95% CI −0.7 to 0.6)). Limits of agreement were wide, particularly for height. Differences were generally random, although routine data tended to underestimate length in taller infants and underestimate weight in lighter infants.

Conclusions Routine data can provide an accurate and feasible method of data collection for research, though wide limits of agreement between data sources may be observed. Differences could be due to methodological issues; but may relate to variability in clinical practice. Continued provision of appropriate training and assessment is essential for health professionals responsible for collecting routine data.

BACKGROUND
Routine data on growth in childhood are widely collected in many countries and used to monitor populations and individuals with respect to health and development.1 2 Increasingly, these data are also used in research. Knowing how these data compare with similar measurements collected in more controlled research settings is important for their use in clinical/public health practice and research.

Use of routine data in research has many advantages; data describing varied outcomes are readily available, they enable inclusion of large samples of the population, and can provide a feasible, cost-effective method to collect information in a way that is more acceptable to patients.1 Use of routine data also offers lower burden to investigators for research conduct and ethical approval. However, there may be drawbacks to using routine data. It may be assumed that they are less reliable and valid than data that have been collected by researchers who have been trained in methods to improve accuracy and reliability and to avoid bias.2 Routine data are not usually collected blindly and may, therefore, be influenced by knowledge of an

What is already known on this topic
- Additional to providing a comprehensive health record, the Personal Child Health Record (PCHR) has the potential to be an excellent source of data for use in research settings.
- Use of routine data for research has many advantages (readily available, large samples, feasible, cost effective), but these data may be less reliable and/or valid.
- Infant height and weight data collected in the early 1990s in a predominantly Causian sample suggest agreement between researcher and PCHR data.

What this study adds
- Although this study found general agreement between Personal Child Health Record (PCHR) data and those collected by researchers (average of 0.5 cm difference), in many instances, wide limits of agreement were observed, with differences in length up to 5 cm in some children.
- Differences could be due to methodological issues (eg, collected on different days); however, they may relate to inaccuracies caused by variability in clinical practice.
- Results indicate that routine data can be an excellent and accurate resource for researchers. But there is a need to ensure health visitors receive regular training and quality assurance to ensure that collection of data relating to height and weight are accurate, for the purposes of clinical practice and for increased use by researchers.
methods is published elsewhere.5

Singleton birth and 28 had twin births. A full account of the

March 2009 were invited to participate in more detailed

characteristic of the city

full BiB cohort recruited 12,453 women comprising 13,776

–

setting. This study aims to determine the degree of agreement

no contemporary data comparing length/height and weight in

using a Seca Leicester height measure (Harlow Healthcare,

using a Seca Leicester height measure (Harlow Healthcare,

original article

methods is published elsewhere.5

singleton birth and 28 had twin births. A full account of the

March 2009 were invited to participate in more detailed

characteristic of the city

full BiB cohort recruited 12,453 women comprising 13,776

–

setting. This study aims to determine the degree of agreement

no contemporary data comparing length/height and weight in

using a Seca Leicester height measure (Harlow Healthcare,

using a Seca Leicester height measure (Harlow Healthcare,

original article

methods is published elsewhere.5

singleton birth and 28 had twin births. A full account of the

March 2009 were invited to participate in more detailed

characteristic of the city

full BiB cohort recruited 12,453 women comprising 13,776

–

setting. This study aims to determine the degree of agreement

no contemporary data comparing length/height and weight in

using a Seca Leicester height measure (Harlow Healthcare,

using a Seca Leicester height measure (Harlow Healthcare,

original article

methods is published elsewhere.5

singleton birth and 28 had twin births. A full account of the

March 2009 were invited to participate in more detailed

characteristic of the city

full BiB cohort recruited 12,453 women comprising 13,776

–

setting. This study aims to determine the degree of agreement

no contemporary data comparing length/height and weight in

using a Seca Leicester height measure (Harlow Healthcare,

using a Seca Leicester height measure (Harlow Healthcare,

original article

methods is published elsewhere.5

singleton birth and 28 had twin births. A full account of the

March 2009 were invited to participate in more detailed

characteristic of the city

full BiB cohort recruited 12,453 women comprising 13,776

–

setting. This study aims to determine the degree of agreement

no contemporary data comparing length/height and weight in

using a Seca Leicester height measure (Harlow Healthcare,

using a Seca Leicester height measure (Harlow Healthcare,
measurement compared to the BiB1000 measurement, and negative coefficients indicate a higher PCHR measurement. We used an approach which could model scale and shape parameters (Generalized Additive Models for Location Scale and Shape) as implemented in the GAMLSS library in R.14 We examined the following mother and baby-related covariates: mother’s ethnicity, mother’s education, infant’s gender, z-score for weight, z-score for height, gestation (<37 weeks or ≥37 weeks) and low birth weight. We allowed the SD parameter of each GAMLSS model to vary with the differences in z-scores for weight plus z-scores for height. Analyses were performed in Stata/IC V.12.1 (StataCorp, College Station, Texas, USA) and GAMLSS in R, V.4.2.8 (R Development Core Team, 2013, London).

**RESULTS**

Figure 1 shows the study sample at each target age. The number of children with BiB1000 and PCHR height and weight measurements, respectively, included at each assessment was: 6 months (n=158 and 560), 12 months (n=101 and 166), 18 months (n=7 and 23), 24 months (n=307 and 434) and 36 months (n=33 and 56). Measurements at 18 and 36 months

---

**Table 1** Mean age, height and weight of BiB1000 and PCHR measurements

<table>
<thead>
<tr>
<th>BiB1000 assessment HEIGHT</th>
<th>Number of children with PCHR height data ±2 months of BiB1000</th>
<th>Mean (SD) age at BiB1000 assessment months</th>
<th>Mean (SD) age at PCHR measurement months</th>
<th>Mean (95% CI) age difference between BiB and PCHR measurements, months</th>
<th>Mean (SD) height, BiB1000 (cm)</th>
<th>Mean (SD) height, PCHR (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>158</td>
<td>6.9 (0.8)</td>
<td>7.7 (0.9)</td>
<td>−0.7 (−0.9 to −0.6)</td>
<td>69.2 (3.3)</td>
<td>69.8 (3.1)</td>
</tr>
<tr>
<td>12 months</td>
<td>101</td>
<td>12.2 (0.8)</td>
<td>11.5 (1.1)</td>
<td>0.7 (0.5 to 0.9)</td>
<td>75.6 (2.9)</td>
<td>75.0 (3.3)</td>
</tr>
<tr>
<td>24 months</td>
<td>307</td>
<td>24.9 (0.7)</td>
<td>24.5 (0.9)</td>
<td>0.5 (0.4 to 0.6)</td>
<td>86.5 (3.4)</td>
<td>86.6 (3.4)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Weight</th>
<th>Mean (SD) weight, BiB1000 (kg)</th>
<th>Mean (SD) weight, PCHR (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>6.7 (0.7)</td>
<td>7.1 (1.2)</td>
</tr>
<tr>
<td>12 months</td>
<td>12.4 (0.8)</td>
<td>11.8 (1.2)</td>
</tr>
<tr>
<td>24 months</td>
<td>25.0 (0.7)</td>
<td>24.5 (0.9)</td>
</tr>
</tbody>
</table>

PCHR, Personal Child Health Record.
were excluded from further analyses due to the small sample sizes; resulting from a lack of routine data available within 2 months of researcher-collected data. Table 1 shows the mean (SD) age, height/length and weight at each assessment. Children tended to be younger at the 6 month BiB1000 assessment and older at the 12 and 24 month assessments compared to PCHR measurements. This table also presents the mean difference and 95% level of agreements for differences in age at which the BiB1000 and PCHR measurements were conducted. Mean differences were less than 1 month at all ages.

There was agreement between data collected from the BiB1000 team and those collected for the PCHR, though this was weaker for height than for weight (table 2; figure 2). Height was somewhat underestimated in routine compared with research data at 6 months, and overestimated at 12 and 24 months. Weight was slightly overestimated in routine compared with research data at all three ages. Limits of agreement were wide, particularly for height (table 2; figure 2). Correlations between differences in routine and researcher-collected z-score data for height and weight measurements indicated that measurement error was non-systematic, with r values of 0.07 at 6 months (i.e., between differences in z-scores for height and differences in z-scores for weight), 0.19 at 12 months and 0.07 at age 24 months. This suggested that differences observed between both sources of data for height were independent from those seen for weight.

Multivariable analysis to explore whether known characteristics (e.g., ethnicity, maternal education, infant gender, gestation, birth weight and height/length and weight) impacted on measurement error showed evidence of systematic bias in relation to mean measurements for height/length (see online supplementary table S1), where routine PCHR data underestimated the infants length by 0.54 cm (95% CI 0.18 to 0.89) more for every additional 1 cm mean height of an infant at 6 months (i.e., PCHR underestimated length in shorter children). The overestimation of height at 12 and 24 months was random with respect to mean height. PCHR height data was significantly lower in Pakistani infants at 6 months compared to Caucasian infants (0.83 cm (95% CI 0.09 to 1.57, p<0.03). Multivariable analyses also showed an impact of gestation on the agreement in weight data at 6 and 12 months, with higher PCHR weight in preterm infants (gestation <37 weeks) at the 6-month measurement (0.82 kg greater (95% CI −1.01 to −0.64, p<0.0001) compared to gestation ≥37 weeks; but lower PCHR weight in preterm infants at 12 months (0.64 kg (95% CI 0.36 to 0.93, p<0.001)). At 12 months, PCHR weight data was greater in the infants of mothers with A level education by 0.18 kg (95% CI −0.31 to −0.04, p=0.014) compared to infants of mothers with ≥5 GCSEs. Other differences in weight measurement appeared random and were not associated with any other characteristics (see online supplementary table S1).

When we repeated the analyses while only including infants within a routine data measurement within 1 month of the research sample, there did not appear to be strong evidence that the agreement was better compared to the main analyses in which we allowed a 2-month difference, though the sample sizes were small (N=38–170 for different analyses). At 6 months, the mean z-score differences were 0.27 for height and −0.02 for weight (n=49 and 296); at 12 months they were −0.19 for height and −0.12 for weight (n=38 and 87), and at 24 months the differences were −0.14 and −0.08 for height and weight, respectively, (n=170 and 264) (results not shown).

**DISCUSSION**

Bland–Altman plots from the current study and those of Howe *et al.*** suggest that routine data collection of height and weight shows little evidence of bias when compared to research data. However, both demonstrated relatively wide limits of agreement around the estimate. So, while population averages appear accurate, there were substantial differences for some individuals, with predicted differences between PCHR and research data ranging between −0.4 to 4.9 cm for height and −1.19–0.98 kg for weight. This may be related to measurement error, but may also be attributable to differences in the dates in which assessments were made. The current study only included PCHR data that were collected within 2 months of the research-collected data (with mean differences all less than 1 month); similar to the methodology applied previously by Howe *et al.* Using this approach, data were normalised for differences in the actual date that the data were collected; however, normalisation relies on the theory that infants track uniformly along a growth curve at a constant rate. In reality, individual variation in growth means that many infants do not always track uniformly along growth chart trajectories. We conducted additional sensitivity analysis to explore this bias; including only data collected within 1 month of each other. Findings suggest similar agreement in mean values but wider levels of agreement; however, the sample size was substantially reduced.

**Table 2** Mean (SD) differences for z-scores of height and weight, and mean differences with 95% levels of agreement for predicted height (cm) and weight (kg) between BiB1000 and PCHR data

<table>
<thead>
<tr>
<th>Assessment age</th>
<th>Mean BiB1000 height z-score (SD)</th>
<th>Mean PCHR height z-score (SD)</th>
<th>Mean height z-score (SD) difference</th>
<th>Mean difference in predicted height (95% level of agreement; mean differences ±1.96× SD) (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>0.49 (1.39)</td>
<td>0.29 (1.16)</td>
<td>0.21 (1.01)</td>
<td>0.46 (−3.99 to 4.91)</td>
</tr>
<tr>
<td>12 months</td>
<td>0.12 (1.16)</td>
<td>0.23 (1.27)</td>
<td>−0.10 (0.85)</td>
<td>−0.25 (−4.50 to 4.00)</td>
</tr>
<tr>
<td>24 months</td>
<td>−0.23 (1.02)</td>
<td>−0.13 (1.02)</td>
<td>−0.10 (0.58)</td>
<td>−0.32 (−4.00 to 3.36)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assessment age</th>
<th>Mean BiB1000 weight z-score (SD)</th>
<th>Mean PCHR weight z-score (SD)</th>
<th>Mean weight z-score (SD) difference</th>
<th>Mean difference in predicted weight (95% level of agreement; mean differences ±1.96× SD) (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 months</td>
<td>−0.11 (1.06)</td>
<td>−0.07 (1.08)</td>
<td>−0.04 (0.37)</td>
<td>−0.04 (−0.67 to 0.59)</td>
</tr>
<tr>
<td>12 months</td>
<td>0.27 (1.07)</td>
<td>0.33 (1.11)</td>
<td>−0.05 (0.44)</td>
<td>−0.06 (−1.10 to 0.98)</td>
</tr>
<tr>
<td>24 months</td>
<td>0.23 (1.02)</td>
<td>0.33 (1.01)</td>
<td>−0.10 (0.35)</td>
<td>−0.14 (−1.19 to 0.91)</td>
</tr>
</tbody>
</table>

PCHR, Personal Child Health Record.
In the BiB1000 protocol, there was a requirement for repeated measures (taken on the same day) to be within 0.5 cm and 0.1 kg of each other. Guidance from the International Fetal and Newborn Growth consortium (Inter-growth-21st) suggest that maximum allowed difference in repeated measures of length should be 0.7 cm, and should be 0.5 kg in measures of weight. Fifty per cent of the BiB1000 data were outside both these parameters. However, these protocols/guidelines refer to measurements conducted on the same child on the same day, and are, therefore, not as relevant to the present study, which compares measurements taken by different people, which are unlikely to take place on the same day.

In this population, routinely collected health visitor data tended to underestimate height for measurements taken when infants were aged 6 months with agreement improving with age. Poorer agreement for data relating to length/height and weight measurements in PCHR has been previously assessed by comparing PCHR data to data collected by researchers as part of the ALSPAC cohort study in the UK using data that were collected almost 20 years ago. In comparison with the BiB data used here, in which the sample are multiethnic (predominately of Pakistani or White British origin), the ALSPAC cohort are a predominantly White British population. Similar patterns in the differences between height and weight z-scores over time were observed, compared to the present study, although mean differences were higher in data presented by Howe et al with the exception of height at 24 months. Similar to the present study, Howe et al observed greatest disagreements in height at younger ages. Unlike Howe et al, the current study was also able to examine any potential influence of ethnicity, and showed a greater likelihood of underestimation of length in Pakistani children at 6 months.

Figure 2 Agreement between PCHR and BiB1000.

In the BiB1000 protocol, there was a requirement for repeated measures (taken on the same day) to be within 0.5 cm and 0.1 kg of each other. Guidance from the International Fetal and Newborn Growth consortium (Inter-growth-21st) suggest that maximum allowed difference in repeated measures of height should be 0.7 cm, and should be 0.5 kg in measures of weight. Fifty per cent of the BiB1000 data were outside both these parameters. However, these protocols/guidelines refer to measurements conducted on the same child on the same day, and are, therefore, not as relevant to the present study, which compares measurements taken by different people, which are unlikely to take place on the same day.

In this population, routinely collected health visitor data tended to underestimate height for measurements taken when infants were aged 6 months with agreement improving with age. Poorer agreement for data relating to length/height and weight measurements in PCHR has been previously assessed by comparing PCHR data to data collected by researchers as part of the ALSPAC cohort study in the UK using data that were collected almost 20 years ago. In comparison with the BiB data used here, in which the sample are multiethnic (predominately of Pakistani or White British origin), the ALSPAC cohort are a predominantly White British population. Similar patterns in the differences between height and weight z-scores over time were observed, compared to the present study, although mean differences were higher in data presented by Howe et al with the exception of height at 24 months. Similar to the present study, Howe et al observed greatest disagreements in height at younger ages. Unlike Howe et al, the current study was also able to examine any potential influence of ethnicity, and showed a greater likelihood of underestimation of length in Pakistani children at 6 months.

Routine data collected for the PCHR is essential in clinical practice, for monitoring, to help provide continuity of care, and as a means to improve communication with parent.
In the current study, we evaluated the reliability of routine anthropometric data collected by health workers for assessing early life growth and its relationship with the risk of obesity. We found that the reliability of routine data varied across different measures, and some measures showed poor agreement with individual cases. This highlights the importance of ensuring that staff are trained and the use of data collection is accurate in measurement. We would advocate using these data for research purposes, with considerations to ensure that the data are collected and used appropriately.

Acknowledgements
We are grateful to all the families who took part in this study, to the midwives for their help in recruiting them, to the paediatricians and health visitors, and to the Born in Bradford team which included interviewers, data managers, laboratory staff, clerical workers, research scientists, volunteers and managers.

Collaborators
The Born in Bradford Childhood Obesity Scientific Group: Amanda Farrin, Helen Ball, Carolyn Summerbell, Sally Barber, Andrew Hill, Neil Small, Pauline Raynor and Rosie McEachan.

Contributors
MB led the design, conduct and writing of the study and manuscript. GS and MM wrote the statistical analysis plan, and cleaned and analysed the data. LF, KT, ES, LDH, MM and DAL all provided expertise in the design and analysis plan and contributed towards interpretation of the findings. NC, DF, MM and JW also contributed to the interpretation of findings, providing additional clinical insight. All authors reviewed and revised the manuscript. All member of the Born in Bradford Childhood Obesity Scientific Group designed and managed the cohort study from which the data were derived and provided direct expertise to the submitted study.

Funding
This work was funded by an NIHR CLaHR implementation grant and an NIHR applied programme grant (RP-PG-0407-10044). This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under the CLaHR programme. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests
All authors had some financial support from an NIHR CLaHR implementation grant and/or an NIHR applied programme grant (RP-PG-0407-10044) for the submitted work, but have had no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work. LDH is funded by a UK Medical Research Council fellowship (G1002375). DAL, KT and LDH work in a unit that receives core funding from the UK Medical Research Council and the University of Bristol (MC_UU_12013/3).

Ethics approval
Bradford Research Ethics Committee (Ref 07/H1302/112).

Provenance and peer review
Not commissioned; externally peer reviewed.

Data sharing statement
Authors have provided a non-identifiable dataset linked to the current analysis, which will be uploaded as a web based file. For further information on sharing of Born in Bradford data, see information provided at http://www.borninbradford.nhs.uk/

Open Access
This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

REFERENCES